CSC 449, HW#2, Kefu Zhu

Problem 1

To compute $\frac{\partial L}{\partial y_i}$, we use the chain rule to break it into two parts $\frac{\partial L}{\partial y_i} = \frac{\partial L}{\partial p_{gt}} \cdot \frac{\partial p_{gt}}{\partial y_i}$

$\because L = \begin{cases} -log(p_{gt}),\ i = gt \\ -log(1 - p_{gt}),\ i \neq gt \end{cases}$ $\quad \therefore \frac{\partial L}{\partial p_{gt}} = \begin{cases} -\frac{1}{p_{gt}},\ i = gt \\ -\frac{1}{1-p_{gt}},\ i \neq gt \end{cases}$

Since $p_{gt} = \frac{e^{y_{gt}}}{\sum_{i=1}^{n} e^{y_i}}$ can be expressed in the form of $f(x) = \frac{g(x)}{h(x)}$, then by the quotient rule, we have

$$f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{|h(x)|^2}$$

Denote $\frac{\partial}{\partial y_i}$ as ∂ and $\sum_{i=1}^{n} e^{y_i}$ as $\sum$ for clear notation

$\therefore \frac{\partial p_{gt}}{\partial y_i} = \frac{\partial\,(e^{y_{gt}}) \cdot \sum\ - e^{y_{gt}} \cdot \partial\,(\sum)}{\sum^2}$

$\because \partial\,(e^{y_{gt}}) = 1, \partial\,(\sum) = e^{y_i}$

$\therefore \frac{\partial p_{gt}}{\partial y_i} = \frac{e^{y_{gt}} \cdot \sum - e^{y_{gt}} \cdot e^{y_{gt}}}{\sum^2} = \frac{e^{y_{gt}}(\sum - e^{y_{gt}})}{\sum^2} = \frac{e^{y_{gt}}}{\sum} \cdot \frac{(\sum - e^{y_{gt}})}{\sum}$

Recall that $p_{gt} = \frac{e^{y_{gt}}}{\sum_{i=1}^{n} e^{y_i}}$, hence $\frac{\partial p_{gt}}{\partial y_i} = p_{gt} \cdot (1 - p_{gt})$

Therefore, we have $\frac{\partial L}{\partial y_i} = \frac{\partial L}{\partial p_{gt}} \cdot \frac{\partial p_{gt}}{\partial y_i} = \begin{cases} -\frac{1}{p_{gt}} \cdot p_{gt} \cdot (1 - p_{gt}) = p_{gt} - 1,\ i = gt \\ -\frac{1}{1-p_{gt}} \cdot p_{gt} \cdot (1 - p_{gt}) = p_{gt}\ ,\ i = gt \end{cases}$

Problem 2

$\because y = Wx + b$

$\therefore$ By the chain rule, we have

$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial W} = \frac{\partial L}{\partial y} \cdot x$

$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial b} = \frac{\partial L}{\partial y} \cdot 1$

$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial x} = \frac{\partial L}{\partial y} \cdot W$

Problem 3

$\because\ y(k,i,j) = \sum_{t=0}^{T-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x(t, i \times s + m, j \times s + n) W_k(t,m,n) + b_k$

$\therefore$ By the chain rule, we have

$$\frac{\partial L}{\partial W_k(t,m,n)} = \frac{\partial L}{\partial y(k,i,j)} \cdot \frac{\partial y(k,i,j)}{\partial W_k(t,m,n)} = \begin{cases} \frac{\partial L}{\partial y(k,i,j)} \cdot x(t, i \times s + m, j \times s + n),\ (m = m, t = t, n = n) \\ 0,\ otherwise \end{cases}$$

$$\frac{\partial L}{\partial b_k} = \frac{\partial L}{\partial y(k,i,j)} \cdot \frac{\partial y(k,i,j)}{\partial b} = \frac{\partial L}{\partial y(k,i,j)} \cdot 1 = \frac{\partial L}{\partial y}$$

$$\frac{\partial L}{\partial x(t,m,n)} = \frac{\partial L}{\partial y(k,i,j)} \cdot \frac{\partial y(k,i,j)}{\partial x(t,m,n)} = \begin{cases} \frac{\partial L}{\partial y(k,i,j)} \cdot W_k(t,m,n),\ (i = 0, j = 0) \\ 0,\ otherwise \end{cases}$$

Problem 4

$\because\ y(c,i,j) = \max_{m=0 \ldots M-1} \max_{n=0 \ldots N-1} x(c, i \times s + m, j \times s + n)$

$\therefore$ By the chain rule, we have

$$\frac{\partial L}{\partial x(t,m,n)} = \frac{\partial L}{\partial y(t,i,j)} \cdot \frac{\partial y(t,i,j)}{\partial x(t,m,n)} = \begin{cases} \frac{\partial L}{\partial y(c,i,j)},\ \ \max_{m=0 \ldots M-1} \max_{n=0 \ldots N-1} x(c, i \times s + m, j \times s + n) = x(t,m,n) \\ 0, otherwise \end{cases}$$